Part I

Markov Chains and Stochastic Sampling

1 Markov Chains and Random Walks on Graphs

1.1 Structure of Finite Markov Chains

We shall only consider Markov chains with a finite, but usually very large, *state space* $S = \{1, \ldots, n\}$.

An *S-valued* (discrete-time) *stochastic process* is a sequence $X_0, X_1, X_2, \ldots$ of S-valued random variables over some probability space Ω, i.e. a sequence of (measurable) maps $X_t : \Omega \to S, t = 0, 1, 2, \ldots$

Such a process is a *Markov chain* if for all $t \geq 0$ and any $i_0, i_1, \ldots, i_{t-1}, i, j \in S$ the following "memoryless" (forgetting) condition holds:

$$\begin{aligned} &\Pr(X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i_{t-1}, X_t = i) \\ &\quad = \quad \Pr(X_{t+1} = j \mid X_t = i). \end{aligned} \tag{1}$$

Consequently, the process can be described completely by giving its *initial distribution (vector)*[1]

$$p^0 = [p_1^0, \ldots, p_n^0] = \left[p_i^0\right]_{i=1}^n, \quad \text{where } p_i^0 = \Pr(X_0 = i)$$

[1]By a somewhat confusing convention, distributions in Markov chain theory are represented as row vectors. We shall be denoting the $1 \times n$ *column* vector with components $p_1, \ldots, p_n$ as $(p_1, \ldots, p_n)$, and the corresponding $n \times 1$ *row* vector as $[p_1, \ldots, p_n] = (p_1, \ldots, p_n)^T$. All vectors shall be column vectors unless otherwise indicated by text or notation.

and its sequence of *transition (probability) matrices*

$$P^{(t)} = \left(p_{ij}^{(t)}\right)_{i,j=1}^{n}, \qquad \text{where } p_{ij}^{(t)} = \Pr(X_t = j \mid X_{t-1} = i).$$

Clearly, by the rule of total probability, the distribution vector at time $t \geq 1$

$$p^{(t)} = [\Pr(X_t = j)]_{j=1}^{n}$$

is obtained from $p^{(t-1)}$ simply by computing for each j:

$$p_j^{(t)} = \sum_{i=1}^{n} p_i^{(t-1)} \cdot p_{ij}^{(t)},$$

or more compactly

$$p^{(t)} = p^{(t-1)} P^{(t)}.$$

Recurring back to the initial distribution, this yields

$$p^{(t)} = p^0 P^{(1)} P^{(2)} \cdots P^{(t)}. \tag{2}$$

If the transition matrix is time-independent, i.e. $P^{(t)} = P$ for all $t \geq 1$, the Markov chain is *homogeneous*, otherwise *inhomogeneous*. We shall be mostly concerned with the homogeneous case, in which formula (2) simplifies to

$$p^{(t)} = p^0 P^t.$$

We shall say in general that a vector $q \in \mathbb{R}^n$ is a *stochastic vector* if it satisfies

$$q_i \geq 0 \quad \forall\, i = 1, \ldots, n \qquad \text{and} \qquad \sum_i q_i = 1.$$

A matrix $Q \in \mathbb{R}^{n \times n}$ is a *stochastic matrix* if all its row vectors are stochastic vectors.

Now let us assume momentarily that for a given homogeneous Markov Chain with transition matrix P and initial probability distribution p^0 there exists a limit distribution $\pi \in [0,1]^n$ such that

$$\lim_{t \to \infty} p^{(t)} = \pi \qquad \text{(in any norm, e.g. coordinatewise).} \tag{3}$$

Then it must be the case that

$$\begin{aligned} \pi &= \lim_{t \to \infty} p^0 P^t = \lim_{t \to \infty} p^0 P^{t+1} \\ &= \left(\lim_{t \to \infty} p^0 P^t\right) P = \pi P. \end{aligned}$$

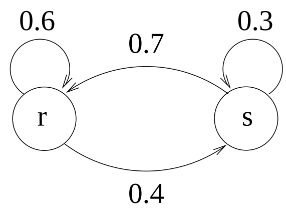

Figure 1: A Markov chain for Helsinki weather.

Thus, any limit distribution satisfying property (3), if such exist, is a left eigenvector of the transition matrix with eigenvalue 1, and can be computed by solving the equation $\pi = \pi P$. Solutions to this equation are called the *equilibrium* or *stationary distributions* of the chain.

Example 1.1 *The weather in Helsinki.* Let us say that tomorrow's weather is conditioned on today's weather as represented in Figure 1 or in the transition matrix:

P	rain	sun
rain	0.6	0.4
sun	0.7	0.3

Then the long-term weather distribution can be determined, in this case uniquely and in fact independent of the initial conditions, by solving

$$
\begin{aligned}
& \pi P = \pi, \quad \sum_i \pi_i = 1 \\
\Leftrightarrow\ & \begin{bmatrix} \pi_r & \pi_s \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \end{bmatrix} = \begin{bmatrix} \pi_r & \pi_s \end{bmatrix}, \quad \pi_r + \pi_s = 1 \\
\Leftrightarrow\ & \begin{cases} \pi_r = 0.6\pi_r + 0.7\pi_s \\ \pi_s = 0.4\pi_r + 0.3\pi_s \end{cases}, \quad \pi_r + \pi_s = 1 \\
\Leftrightarrow\ & \begin{cases} \pi_r = 0.64 \\ \pi_s = 0.36 \end{cases}
\end{aligned}
$$

Every finite Markov chain has at least one stationary distribution, but as the following examples show, this need not be unique, and even if it is, then the chain does not need to converge towards it in the sense of equation (3).

Example 1.2 *A reducible Markov chain.* Consider the chain represented in Figure 2. Clearly any distribution $p = [p_1 \quad p_2]$ is stationary for this chain. The

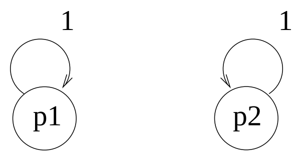

Figure 2: A reducible Markov chain.

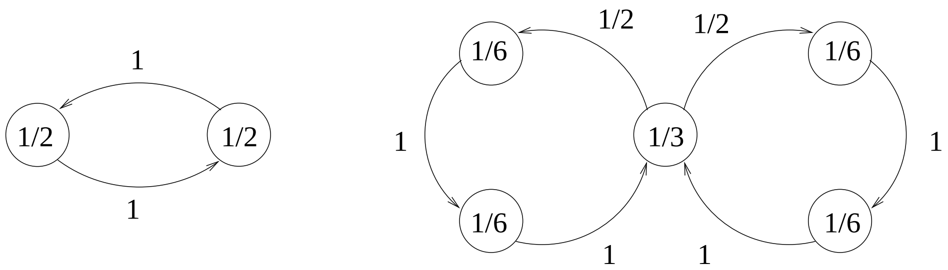

Figure 3: Periodic Markov chains.

underlying cause for the existence of several stationary distributions is that the chain is *reducible*, meaning that it consists of several “noncommunicating” components. (Precise definitions are given below.)

Any irreducible (“fully communicating”) chain has a unique stationary distribution, but this does not yet guarantee convergence in the sense of equation (3).

Example 1.3 *Periodic Markov chains.* Consider the chains represented in Figure 3. These chains are *periodic*, with periods 2 and 3. While they do have unique stationary distributions indicated in the figure, they only converge to those distributions from the corresponding initial distributions; otherwise probability mass “cycles” through each chain.

So when is a unique stationary limit distribution guaranteed? The brief answer is as follows.

Consider a finite, homogeneous Markov chain with state set S and transition matrix P. The chain is:

(i) *irreducible*, if any state can be reached from any other state with positive probability, i.e.

$$\forall\, i, j \in S \quad \exists\, t \geq 0 \;:\; P^t_{ij} > 0;$$

(ii) *aperiodic* if for any state $i \in S$ the greatest common divisor of its possible recurrence times is 1, i.e. denoting

$$N_i = \{t \geq 1 \mid P^t_{ii} > 0\}$$

we have $\gcd(N_i) = 1, \quad \forall\, i \in S$.

Theorem (Markov Chain Convergence) *A finite homogeneous Markov chain that is irreducible and aperiodic has a unique stationary distribution* π, *and the chain will converge towards this distribution from any initial distribution* p^0 *in the sense of Equation (3).* $\square$

Irreducible and aperiodic chains are also called *regular* or *ergodic*.

We shall prove this important theorem below, establishing first the existence and uniqueness of the stationary distribution, and then convergence. Before going into the proof, let us nevertheless first look into the structure of arbitrary, possibly nonregular, finite Markov chains somewhat more closely.

Let the finite state space be S and the homogeneous transition matrix be P.

A set of states $C \subseteq S, C \neq \varnothing$ is *closed* or *invariant*, if $p_{ij} = 0 \quad \forall\, i \in C, j \notin C$.

A singleton closed state is *absorbing* (i.e. $p_{ii} = 1$).

A chain is *irreducible* if S is the only closed set of states. (This definition can be seen to be equivalent to the one given earlier.)

Lemma 1.1 *Every closed set contains a **minimal** closed set as a subset.* $\square$

State j is *reachable* from state i, denoted $i \to j$, if $P^t_{ij} > 0$ for some $t \geq 0$.

States $i, j \in S$ *communicate*, denoted $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Lemma 1.2 *The communication relation "$\leftrightarrow$" is an equivalence relation. All the minimal closed sets of the chain are equivalence classes with respect to "$\leftrightarrow$". The chain is irreducible if and only if all its states communicate.* $\square$

States which do not belong to any of the minimal closed subsets are called *transient*.

One may thus partition the chain into equivalence class with respect to "$\leftrightarrow$". Each class is either a minimal closed set or consists of transient states. This is illustrated in Figure 4. By "reducing" the chain in this way one obtains a DAG-like structure, with the minimal closed sets as leaves and the transient components as internal nodes. (Actually a "forest" if the chain is disconnected.) An irreducible chain of course reduces to a single node.

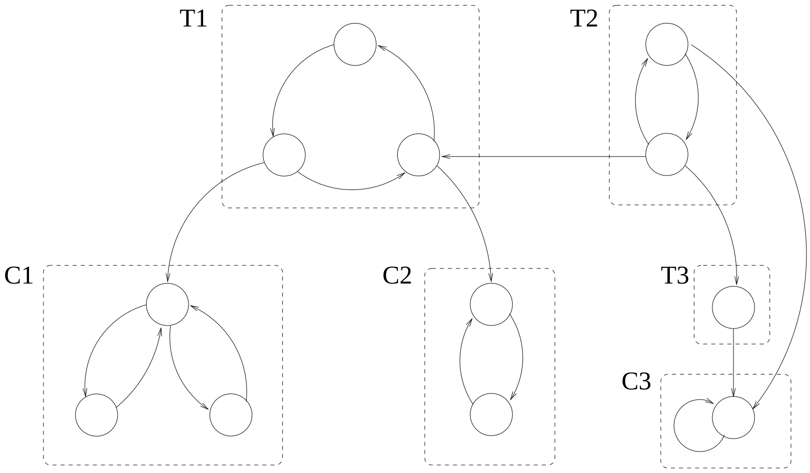

Figure 4: Partitioning of a Markov chain into communicating classes.

The *period* of state $i \in S$ is

$$\gcd\{\underbrace{t \geq 1 \mid P_{ii}^t > 0}_{N_i}\}.$$

A state with period 1 is *aperiodic*.

Lemma 1.3 *Two communicating states have the same period. Hence, every component of the "$\leftrightarrow$" relation has a uniquely determined period.* $\square$

Define the *first hit* (or *first passage*) probabilities for states $i \to j$ and $t \geq 1$ as:

$$f_{ij}^{(t)} = \Pr(X_1 \neq j, X_2 \neq j, \ldots, X_{t-1} \neq j, X_t = j \mid X_0 = i),$$

and the *hitting* (or *passage*) *probability* for $i \to j$ as

$$f_{ij}^* = \sum_{t \geq 1} f_{ij}^{(t)}.$$

Then the *expected hitting* (*passage*) *time* for $i \to j$ is

$$\mu_{ij} = \begin{cases} \sum_{t \geq 1} t f_{ij}^{(t)}, & \text{if } f_{ij}^* = 1; \\ \infty & \text{if } f_{ij}^* < 1 \end{cases}$$

For $i = j$, μ_{ii} is called the *expected return time*, and often denoted simply μ_i.

State $i \in S$ is *recurrent* (or *persistent*) if $f_{ii}^* = 1$, otherwise it is *transient*. (In infinite Markov chains the recurrent states are further divided into *positive recurrent* with $\mu_i < \infty$ and *null recurrent* with $\mu_i = \infty$, but the latter case does not occur in finite Markov chains and thus need not concern us here.)

The following theorem provides an important characterisation of the recurrent states.

Notation: $P^k = \left(p_{ij}^{(k)}\right)_{i,j=1}^{n}$.

Theorem 1.4 *State $i \in S$ is recurrent if and only if $\sum_{k\geq 0} p_{ii}^{(k)} = \infty$. Correspondingly, $i \in S$ is transient if and only if $\sum_{k\geq 0} p_{ii}^{(k)} < \infty$.*

Proof. Recall the relevant definitions:

$$\begin{aligned} p_{ii}^{(k)} &= \Pr(X_k = i \mid X_0 = i), \\ f_{ii}^{(t)} &= \Pr(X_1 \neq i, \ldots, X_{t-1} \neq i, X_t = i \mid X_0 = i). \end{aligned}$$

Then it is fairly clear that

$$p_{ii}^{(k)} = \sum_{t=1}^{k} f_{ii}^{(t)} p_{ii}^{(k-t)} = \sum_{t=0}^{k-1} f_{ii}^{(k-t)} p_{ii}^{(t)}.$$

Consequently, for any K:

$$\begin{aligned} \sum_{k=1}^{K} p_{ii}^{(k)} &= \sum_{k=1}^{K} \sum_{t=0}^{k-1} f_{ii}^{(k-t)} p_{ii}^{(t)} \\ &= \sum_{t=0}^{K-1} p_{ii}^{(t)} \sum_{k=t+1}^{K} f_{ii}^{(k-t)} \\ &\leq \sum_{t=0}^{K} p_{ii}^{(t)} f_{ii}^{*} \\ &= \left(1 + \sum_{t=1}^{K} p_{ii}^{(t)}\right) f_{ii}^{*} \end{aligned}$$

Since K was arbitrary, we obtain:

$$(1 - f_{ii}^{*}) \sum_{k=1}^{\infty} p_{ii}^{(k)} \leq f_{ii}^{*}.$$

Now if $i \in S$ is transient, i.e. $f_{ii}^{*} < 1$, then

$$\sum_{k\geq 1} p_{ii}^{(k)} \leq \frac{f_{ii}^{*}}{1 - f_{ii}^{*}} < \infty.$$

Conversely, assume that $i \in S$ is recurrent, i.e. $f_{ii}^* = 1$. Now one can see that

$$\begin{aligned}\Pr(X_t = i \text{ for at least two } t \geq 1 \mid X_0 = i) &= \sum_{t,t' \geq 1} f_{ii}^{(t)} f_{ii}^{(t')} = \left(\sum_{t \geq 1} f_{ii}^{(t)} \right)^2 \\ &= (f_{ii}^*)^2 = 1,\end{aligned}$$

and by induction that

$$\Pr(X_t = i \text{ for at least } s \text{ times } | X_0 = i) = (f_{ii}^*)^s = 1.$$

Consequently,

$$P_{kk}^{\infty} \triangleq \Pr(X_k = i \text{ infinitely often} \mid X_0 = i) = \lim_{s \to \infty} (f_{ii}^*)^s = 1.$$

However, if $\sum_{k \geq 0} p_{ii}^{(k)} < \infty$, then by the Borel-Cantelli lemma (see below) it should be the case that $p_{kk}^{\infty} = 0$.

Thus it follows that if $f_{ii}^* = 1$, then also $\sum_{k \geq 0} p_{ii}^{(k)} = \infty$. □

Lemma (Borel-Cantelli, "easy case") *Let $A_0, A_1, \ldots$ be events, and A the event "infinitely many of the A_k occur". Then*

$$\sum_{k \geq 0} \Pr(A_k) < \infty \Rightarrow \Pr(A) = 0.$$

Proof. Clearly

$$A = \bigcap_{m \geq 0} \bigcup_{k \geq m} A_k.$$

Thus for all $m \geq 0$,

$$\Pr(A) \leq \Pr\left(\bigcup_{k \geq m} A_k \right) \leq \sum_{k \geq m} \Pr(A_k) \to 0 \text{ as } m \to \infty,$$

assuming the sum $\sum_{k \geq 0} \Pr(A_k)$ converges. □

Let $C_1, \ldots, C_m \subseteq S$ be the minimal closed sets of a finite Markov chain, and $T \triangleq S \setminus (C_1 \cup \cdots \cup C_m)$.

Theorem 1.5 *(i) Any state $i \in C_r$, for some $r = 1, \ldots, m$, is recurrent.*
(ii) Any state $i \in T$ is transient.

Proof. (i) Assume $i \in C$, C minimal closed subset of S. Then for any $k \geq 1$,

$$\sum_{j \in S} p_{ij}^{(k)} = \sum_{j \in C} p_{ij}^{(k)} = 1,$$

because C is closed and P is a stochastic matrix. Consequently,

$$\sum_{k \geq 0} \sum_{j \in C} p_{ij}^{(k)} = \infty,$$

and because C is finite, there must be some $j_0 \in C$ such that

$$\sum_{k \geq 0} p_{ij_0}^{(k)} = \infty.$$

Since $j_0 \leftrightarrow i$, there is some $k_0 \geq 0$ such that $p_{j_0 i}^{(k_0)} = p_0 > 0$. But then

$$\sum_{k \geq 0} p_{ii}^{(k)} \geq \sum_{k \geq k_0} p_{ij_0}^{(k-k_0)} p_{j_0 i}^{(k_0)} = \left(\sum_{k \geq k_0} p_{ij_0}^{(k-k_0)} \right) \cdot p_0 = \infty.$$

By Theorem 1.4 i is thus recurrent.

(ii) Denote $C = C_1 \cup \cdots \cup C_m$. Since for any $j \in Y$ the set $\{l \in S \mid j \rightarrow l\}$ is closed, it must intersect C; thus for any $j \in T$ there is some $k \geq 1$ such that

$$p_{jC}^{(k)} \triangleq \sum_{l \in C} p_{jl}^{(k)} > 0.$$

Since T is finite, we may find a $k_0 \geq 1$ such that for any $j \in T$, $p_{jC}^{(k_0)} = p > 0$. Then one may easily compute that for any $i \in T$,

$$p_{iT}^{(k_0)} \leq 1 - p,\ p_{iT}^{(2k_0)} \leq (1-p)^2,\ p_{iT}^{(3k_0)} \leq (1-p)^3, \text{ etc.}$$

and so

$$\sum_{k \geq 1} p_{ii}^{(k)} \leq \sum_{k \geq 1} p_{iT}^{(k)} \leq \sum_{r \geq 0} k_0 p_{iT}^{(rk_0)} \leq k_0 \sum_{r \geq 0} (1-p)^r < \infty.$$

By Theorem 1.4, i is thus transient. □

1.2 Existence and Uniqueness of Stationary Distribution

A matrix $A \in \mathbb{R}^{n \times n}$ is